28 February 1998

A Multi-Method Exploration of Crime Hot-Spots: Summary of Findings Utilizing the IDRISI Software Package

By
Doug Williamson
Geography Dept, Hunter College and
Center for Applied Studies of the Environment, City University of New York

Prepared for the 35th Annual Meeting of the Academy of Criminal Justice Sciences:
"*Crime Control Policies: Political Rhetoric or Empirical Reality*"

Introduction

According to a recent directory compiled by *GIS World*, there were approximately 440 GIS software titles available in 1997. While a large number of these are not applicable to Crime Mapping, many are. Crime mapping unlike most GIS (Geographic Information Systems) applications has a clear objective: to identify and monitor hotspots of criminal activity. Hotspots, as defined by Block (1995), are "the densest clusters of incidents on a map". Until now, there has been no systematic comparison of the various tools used to identify hotspots. This analysis will attempt carry out such a comparison between eleven hotspot identification tools. Each of the tools will be used on a common data set containing burglaries and street robberies in Baltimore County from November 1, 1996 through November 31, 1997. The concept of such a systematic analysis was put forth by the staff of the Crime Mapping Research Center of the National Institute of Justice. This paper will focus on one such tool, namely, IDRISI from the Clark Labs for Cartographic Technology and Geographic Analysis at Clark University (Worcester, Massachusetts). It should be noted that until recently (January 1998) the author never worked with IDRISI. Any remarks should considered to be from a novice, not an expert user.

According to the creators of IDRISI, "[It] is designed to be easy to use, yet provide professional-level GIS, Image Processing and Spatial Statistics analytical capability on both DOS- and Windows-based personal computers. It is intended to be affordable to all levels of users, and to run on the most basic of common computer platforms… IDRISI for Windows, first released in 1995, added a graphical user interface, flexible cartographic composition facilities, and an integrated database management system to IDRISI's extensive analytical toolkit."

Using IDRISI

IDRISI, like most modern software packages, is Windows based. Therefore, anyone familiar with the Windows environment will be able to get acclimated fairly easily. However, IDRISI was originally a DOS based program and only made the transition to Windows in 1995. Because of this a lot of the recent functionality associated with Windows 95 and Windows NT has not yet been implemented. For example, file names can only be eight characters long. Related to this is the usage of files and directories. In IDRISI, the utilities to open or access files are rather clumsy. Either the environment must be set to a directory where *all* the files reside, or it must be continuously changed to the location where files are. The creators of IDRISI boast of a new feature, which allows the user to 'double-click' in a filename box to get a list of valid files in that directory. This is useful, except in the instance when the needed files are in another directory, which may very often be the case.

Another practical aspect of the software is the import/export utility within IDRISI. This feature is quite impressive compared to many other software packages. It provides the ability to import and export a variety of software specific formats including ArcView shapefiles, Arc/Info coverages, MapInfo .mif files, Atlas GIS .bna files, and Autocad DXF files. It also allows for image conversion including ERDAS image files, GRASS image files, BIL (band-interleaved by line) image files, BIP (band-interleaved by pixel) image files, Windows bitmap image files and TIFF (tag image file format) images. Also, useful is IDRISI's ability to import/export United States government formats such as DLG (digital line graph), which may be useful for certain law enforcement applications and DEM (digital elevation model).

The import/export utilities provided dramatically reduced the time spent during this analysis. It has been estimated that for any given GIS project, at least 70 percent of the time and resources will be devoted to data input and verification. Very often, a user may have to go through several perturbations before data can finally be used in a software package. In this analysis, data were provided in two formats, ArcView and MapInfo. It was known from the beginning that data would have to be transformed into some projection other than the geographic latitude/longitude format they were provided in. IDRISI was new, so it was unclear as to how easy this process would be once the data were in IDRISI's native format. Because of this, it was decided that the MapInfo data would be used. The only reason for this decision was ease of operation. MapInfo allows the user to save data in a different projection, while ArcView does not. As it turns out re-projecting data is a straightforward procedure in IDRISI. Once the data were successfully imported and opened in IDRISI, the analysis was relatively easy.

As was stated earlier, one of the main functions of IDRISI is image processing. Because of this, IDRISI, for all intents and purposes, is a raster based GIS. Virtually all the algorithms within IDRISI operate on raster data. The quadrat function is no exception. With that in mind, the data needed to go through one final conversion step before the analysis could take place. The vector file containing the point data needed to be converted to a raster image. This was achieved by first creating an empty raster grid with no data, then, by using the vector to raster utilities, the frequency of events is assigned to the corresponding grid cell. The extent of the grid was defined by the extent of the original vector file, while the resolution of the grid was determined by the square root of the areal extent divided by the number of incidents. This number is analogous to the unit area per average number of events.

The last step in the analysis is map generation. Considering this is usually the end result or final product for delivery and therefore the most important output, IDRISI's functionality is limited to say the least. The Map Composer module is rather crude and clumsy for a 'professional-level' GIS. Although the composer allows you to add all the necessary map elements (scale bar, north arrow, etc), it is limited in its ability to manipulate these elements. Unlike other mapping programs or graphics packages, the user can not just simply highlight the element to be changed and alter it. The elements must be defined when the first layer is opened, and only certain modifications, repositioning for example, can be done later. Symbolism is limited to basic symbols in IDRISI. For points, the user is limited to circles or squares of varying color and size. Lines are restricted to single lines of varying colors, widths and patterns (dashed, dotted etc.). There are no double line options for highways or streets. Polygons are limited to changes in color or pattern. Outline widths can not be changed.

Overview of Quadrat Analysis

Hotspot analysis is unique in that it deals primarily with point patterns. That is to say, incidents of crime can be depicted on maps as points, unique geographic features that have no-dimension in space. There are a variety of spatial statistical methods that can be used to explain point patterns, kernel estimation, nearest neighbor tests and the k-function to name a few. Another common method, albeit rudimentary in sophistication is quadrat analysis. It is this method, as implemented by IDRISI that will be the focus of analysis described below.

In its simplest terms, quadrat analysis summarizes a point pattern by partitioning a study area into subregions of equal area, called quadrats, and counts of events are computed for each quadrat. Then, if desired, this can be converted to a measure of intensity by dividing the number of incidents by the geographical area of the quadrat. This can give some indication of how the density of crime varies over the study area. Essentially, the algorithm transforms the original point pattern into a set of area data. The problem with this is that although a global indication of density may be given, much of the original spatial detail (information) is lost through aggregation. According to Bailey and Gatrell (1997), "As quadrats are made smaller to retain more spatial information we get a very high variability in quadrat counts; this finally degenerates into a mosaic with many empty quadrats, making meaningful interpretation impossible."

Once the counts per quadrat have been calculated, a statistical test for spatial randomness can be performed to assess clustering in the point pattern. Assuming the point pattern follows a Poisson distribution, under complete spatial randomness, the mean and the variance of the counts would be equal. This can be tested using the index of dispersion (the variance of the counts / the mean of the counts) and the index of cluster size or ICS ((the variance of the counts / the mean of the counts) –1). In general, "Note that for CSR [complete spatial randomness], E(ICS) = 0 [the expected value of the index of cluster size equals zero], and if ICS > 0 then clustering is implied ('number of extra events'), whilst if ICS < 0 regularity is implied ('deficiency in events')" (Bailey and Gatrell, 1997).

Results from IDRISI

Statistical analyses are quite easy to perform in IDRISI. In the case of quadrat analysis, the user tells IDRISI which image is to be analyzed and IDRISI executes the function. The results from the quadrat analysis indicate that the spatial point pattern of burglaries and robberies in the entire study area (all of Baltimore County) and the focus area (Southwest Baltimore County) are clustered globally. The operative word here is globally. The results from the analysis only give an indication over the entire region being examined. It does not tell the analyst exactly where the areas of highest density are or how significant the highest areas may be. This information can, however, be 'interpreted' from the maps, based on the numbers of events per quadrat

The results below are the actual output from IDRISI, and show the relative amounts of clustering for each set of data; all burglaries, all robberies, southwest burglaries, southwest robberies.

ALL BURGLARIES:

```
n                     :     38683040
mean (density)        :     0.000161
variance              :     0.000255

Variance/Mean Ratio   :     1.584821
t                     :  2571.982910
df                    :     38683039

Significance level    :       <0.001


Variance/mean ratio values significantly
greater than 1 suggest a clustered pattern
```

ALL ROBBERIES:

```
n                     :      1438610
mean (density)        :     0.000849
variance              :     0.001469

Variance/Mean Ratio   :     1.729103
t                     :   618.365173
df                    :      1438609

Significance level    :       <0.001


Variance/mean ratio values significantly
greater than 1 suggest a clustered pattern
```

BURGLARIES (Southwest):

```
n                    :      3809576
mean (density)       :     0.000512
variance             :     0.000789

Variance/Mean Ratio  :     1.539724
t                    :   744.894897
df                   :      3809575

Significance level   :       <0.001


Variance/mean ratio values significantly
greater than 1 suggest a clustered pattern
```

ROBBERIES (Southwest):

```
n                    :       221325
mean (density)       :     0.002124
variance             :     0.003764

Variance/Mean Ratio  :     1.772352
t                    :   256.929993
df                   :       221324

Significance level   :       <0.001


Variance/mean ratio values significantly
greater than 1 suggest a clustered pattern
```

From these outputs, it can be seen that each of the four sets of data is clustered. In each case, the index of cluster size (variance/mean ratio – 1) is greater then zero, which, according to Bailey and Gatrell, indicates clustering. Also, robberies are more clustered than burglaries, since they have a larger index of cluster size. It should be noted that the test statistic, t, is extremely high for each data set. At first glance, this would indicate extreme significance, but upon closer inspection, this may not be the case. In all four cases, the degrees of freedom, df, is quite high, over 200,000. This is a function of the size of the study area and the number of rows and columns used in creating the grid for the quadrat analysis. If a large number of rows and/or columns is used, then a high number of degrees of freedom will result.

The maps also reveal some interesting patterns. Each map shows a certain amount of clustering, particularly for robberies. Each map is unique in the fact that they are based on different size grid cells, and therefore has different numbers of rows and columns. The areal extent for the main study area (Baltimore County) is the same for burglaries and robberies. The extent is also the same burglaries and robberies in the focus area (Southwest Baltimore County). However, since each subset of the data has a different frequency, the cell sizes are different. This can be clearly seen when comparing burglaries and robberies for the entire study area (Figure 1 and Figure 2). There are 6219 burglaries and only 1222 robberies dispersed over the same size area. Because of this, the cell size will be much smaller for burglaries than for robberies, 10 meters compared to 52 meters. Figure 3 and Figure 4 show the clustering in the focus area. Robberies appear to be quite clustered in several areas. Again, these maps reflect the different grid cell sizes used, which will be different depending on the number of events. The cell size can be left constant, although this would be an arbitrary decision, not based on any statistical method. It should be noted that the maps are not the 'original' frequency maps created in the vector to raster conversion. These maps have been filtered using a 7 x 7 mean filter to enhance the patterns. Mean filters are used to generalize an image. In essence, the filter moves a 7 x 7 (grid cells) window across the image. The center cell is assigned the average value of the 48 neighboring cells. The resulting map is a spatially smoothed image, which estimates how the intensity of incidents varies across the study area.

Practical Aspects of IDRISI

Because IDRISI is a raster based GIS, it is rather limited in its analytic capabilities of spatial point patterns, which are inherently vector features. This is not to say IDRISI has no use in Crime Mapping, but merely limited use. Spatial statistics on raster data is, in general, are more tedious than on vector data, primarily due to the large size of the files themselves. Also, when a description of a point pattern is the goal, raster file structures are not usually desired, since much of the spatial information is lost. However, many types of analyses are simplified in raster GIS. Other analytic features of the software include tests for spatial autocorrelation, time-series analysis and map calculations. IDRISI, however, is more suited to fields that require the use of raster data or remotely sensed data such as natural resources or environmental monitoring.

IDRISI can however, be used in conjunction with other software packages. As was discussed earlier, import and export to other mapping and GIS packages is quite easy. Also, files can be exported to desktop publishing packages as either images (bmp, tiff, etc.) or vectors (dxf and wmf). Clearly this would be an improvement over IDRISI's Map Composer module.

Although not done here, other kinds of spatial analyses are available for crime data. It is unlikely that a crime analyst would opt for this, but the options do exist. One example of this would be a test for autocorrelation. This would tell the analyst how similar the crime rates in areas are relative to those in neighboring areas. The user would expect fairly high values of spatial autocorrelation, since one would not expect neighboring grid cells to be dissimilar. Another extension to the analysis would be time-series analysis, which at first glance appears to be relatively simple in IDRISI. The analyst could get a measure of how much values changed over time, either from one period to the next or over a series of periods.

Flexibility of IDRISI

Compared to some other GIS packages, IDRISI is relatively inflexible in its operation. As was described above, map creation can be an arduous process. IDRISI does provide some symbol sets that can be used and/or altered, but these are primarily suited for raster images. The vector symbols sets, almost invariably need to be altered. This can be a time-consuming process, usually resulting from trial and error.

Another instance of IDRISI's rigidity is the process of grid creation. The user has two options, either copy the parameters from another grid or enter their own parameters. In the case of this analysis, as would probably be the case in most analyses, there were no prior grids to copy from, so all the parameters had to be entered. The parameters include: number of columns, number of rows, X min, X max, Y min, and Y max. For most analysis, determining the X, Y values should not be a problem, but calculating the number of rows and columns from this may be, especially if the desired cell size is unknown.

One bright spot on the issue of flexibility is IDRISI's import/export capabilities. IDRISI can work with files from a variety of software and can create equally as many to use elsewhere. This may be particularly useful to compensate for its shortcomings in map creation.

Recommendations/Conclusions

The use of IDRISI may facilitate the statistical analysis of crime point patterns. It includes some statistical algorithms, like quadrat analysis, that are not available in many other GIS packages. Quadrat Analysis can give the analyst a global idea of how clustered a set of points may or may not be over a study area. IDRISI can also create smoothed maps of the crime data, which will help reinforce, or reveal patterns in the data that were previously unknown. Although the mapping capabilities are somewhat limited, IDRISI is very 'friendly'. Because of this, users should have little trouble going back and forth to other software applications. IDRISI is primarily a raster based GIS, and because of this, it is better suited to the analysis of gridded data, such as images, or spatially continuous data i.e., terrain. However, it does support vector data, similar to the crime data analyzed here, but this is not IDRISI's strong point. Certain operations must be performed on vector data, for crime patterns to be analyzed with any utility.

REFERENCES

Bailey, T. and Gatrell, A., (1997) Interactive Spatial Data Analysis, Longman Scientific and Technical, Essex, England.

Block, C., (1995) "STAC Hot-Spot Areas: A Statistical Tool for Law Enforcement Decisions" in Crime Analysis Through Crime Mapping, eds., Block, C., Dabdoub, M and Fregly, S., Police Executive Research Forum, Washington, D.C.

"IDRISI GIS" (world wide web site), http://www.idrisi.clarku.edu/, 1998.

Smoothed Image of Burglaries in Baltimore County

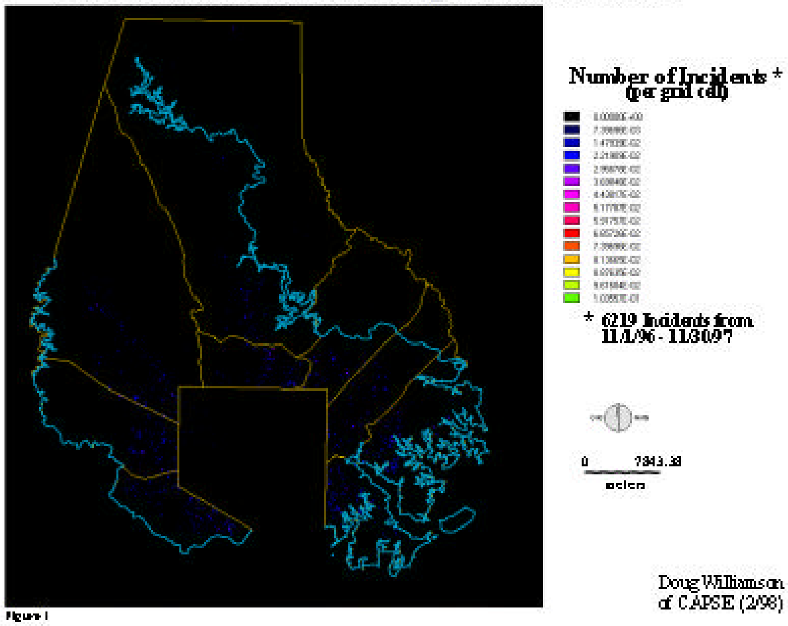


Figure 1

Smoothed Image of Burglaries in Southwest Baltimore County

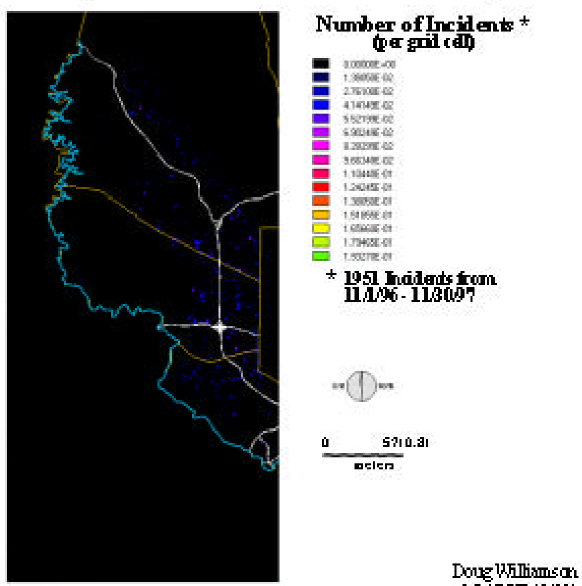
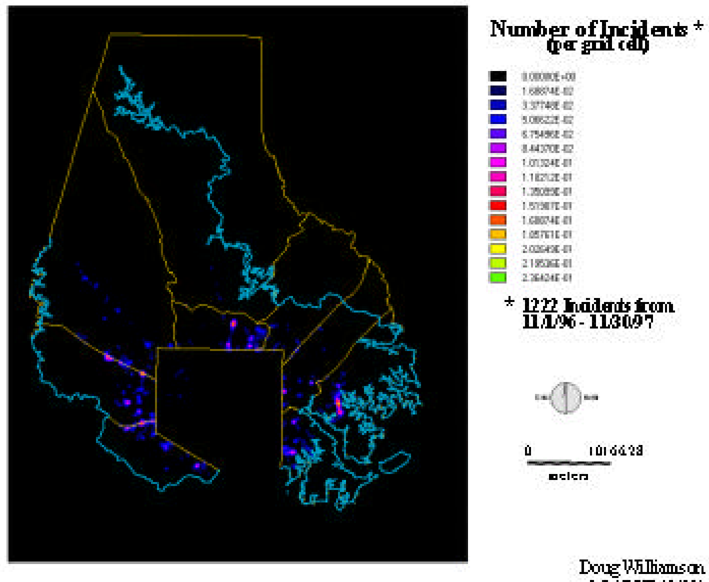


Figure 2

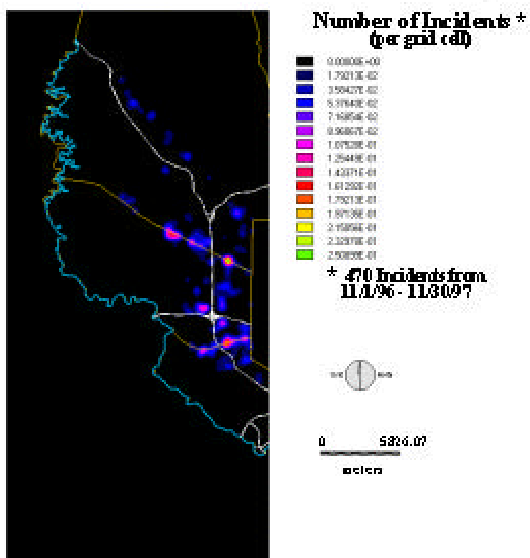
Smoothed Image of Robberies in Baltimore County



Doug Williamson of CAPSE (2/98)

Figure 3

Smoothed Image of Robberies in Southwest Baltimore County



Doug Williamson of CAPSE (2/98)

Figure 4